

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-015491

(43)Date of publication of application : 22.01.1999

(51)Int.Cl.

G10L 3/00

G10L 9/00

G10L 9/18

(21)Application number : 10-163354

(71)Applicant : DIGITAL EQUIP CORP <DEC>

(22)Date of filing : 11.06.1998

(72)Inventor : EBERMAN BRIAN S
MORENO PEDRO J

(30)Priority

Priority number : 97 876601

Priority date : 16.06.1997

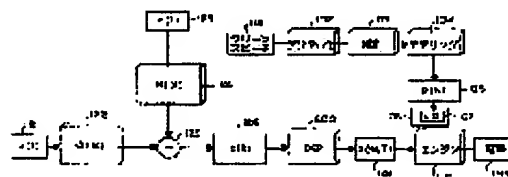
Priority country : US

(54) ENVIRONMENTALLY COMPENSATED METHOD OF PROCESSING SPEECH

(57)Abstract:

PROBLEM TO BE SOLVED: To compensate digitized speech signal with data derived from an acoustic environment by using a clean speech signal without distortion.

SOLUTION: A 1st feature vector representing a clean speech signal 101 is stored in a vector code book 106. A 2nd vector is determined to a dirty speech signal 126 containing noise and distortion 123 parameterized by environmental noise and distortion parameters Q, H, Sn. The noise and distortion parameters are estimated from the 2nd vector. By using the estimated parameters, a 3rd vector is estimated. The 3rd vector is applied to the 2nd vector to form a corrected vector, and by statistically comparing this corrected vector with the 1st vector, it is possible to identify the 1st vector most similar to the corrected vector. Thus, successive data speech signals 126 are compensated by using the estimated values of the environmental noise and distortion parameters Q, H, Sn.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Number of appeal against examiner's decision
of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japan Patent Office

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平11-15491

(43)公開日 平成11年(1999) 1月22日

(51)Int.Cl. ⁸	識別記号	F I
G 1 0 L 3/00	5 2 1	G 1 0 L 3/00
9/00		9/00
9/18		9/18
		5 2 1 L
		F
		E

審査請求 未請求 請求項の数 5 O L (全 11 頁)

(21)出願番号 特願平10-163354

(22)出願日 平成10年(1998) 6月11日

(31)優先権主張番号 08/876601

(32)優先日 1997年6月16日

(33)優先権主張国 米国 (US)

(71)出願人 590002873

ディジタル イクイブメント コーポレイ
ションアメリカ合衆国 テキサス州 77070-
2698 ヒューストン エス. エイチ. 249
-20555

(72)発明者 ブライアン エス エイパーマン

アメリカ合衆国 マサチューセッツ州
02144サマーヴィル ウィロー アベニュー
264

(74)代理人 弁理士 中村 稔 (外6名)

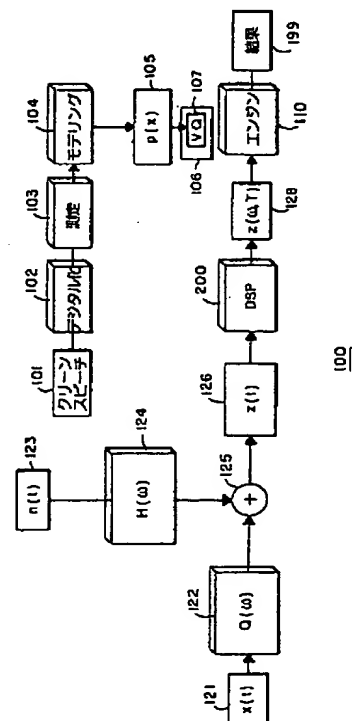
最終頁に続く

(54)【発明の名称】 環境的に補償されたスピーチ処理方法

(57)【要約】

【課題】 スピーチ信号が発生されて伝達される音響環境から導出したデータでデジタルスピーチ信号を補償する方法を提供する。

【解決手段】 スピーチ信号を処理するコンピュータ化された方法において、クリーンスピーチ信号を表す第1ベクトルがベクトルコードブックに記憶される。第2ベクトルは、ダーティスピーチ信号から決定される。第2ベクトルからノイズ及び歪パラメータが推定される。第3ベクトルは、その推定されたノイズ及び歪パラメータに基づいて予想される。第3ベクトルを用いて、第1ベクトルが修正される。次いで、第3ベクトルを第2ベクトルに適用し、修正されたベクトルを発生することができる。修正されたベクトルと第1ベクトルとを比較し、修正されたベクトルに類似する第1ベクトルを識別することができる。



【特許請求の範囲】

【請求項1】 歪のある「ダーティ」信号と称するスピーチ信号を処理するためのコンピュータ化された方法であって、歪のないスピーチ信号は、「クリーン」スピーチ信号と称し、上記方法は、クリーンスピーチ信号を表す第1ベクトルをベクトルコードブックに記憶し、ダーティスピーチ信号から第2ベクトルを決定し、第2ベクトルから環境パラメータを推定し、第1ベクトルを修正するために上記推定された環境パラメータに基づいて第3ベクトルを予想し、第3ベクトルを第2ベクトルに適用して、修正されたベクトルを発生し、そして上記修正されたベクトルと第1ベクトルとを比較して、上記修正されたベクトルに類似した第1ベクトルを識別する、という段階を備えたことを特徴とする方法。

【請求項2】 特定の修正されたベクトルと、それに対応する第1ベクトルとの間の距離を決定し、この距離は、第1ベクトルが上記修正されたベクトルに類似する見込みを表し、更に、特定の修正されたベクトルがそれに対応する第1ベクトルに類似する見込みを最大にする段階を含む請求項1に記載の方法。

【請求項3】 上記比較段階は、統計学的な比較を使用し、この統計学的な比較は、最小平均平方エラーに基づく請求項1に記載の方法。

【請求項4】 上記第1ベクトルは、クリーンスピーチの音素を表し、上記比較段階は、スピーチ認識を行うためにダーティスピーチの内容を決定する請求項1に記載の方法。

【請求項5】 上記第1ベクトルは、既知の話し手のクリーンスピーチのモデルを表し、上記比較段階は、ダーティスピーチ信号を発生する未知の話し手の認識を決定する請求項1に記載の方法。

【発明の詳細な説明】**【0001】**

【発明の属する技術分野】本発明は、一般に、スピーチ処理に係り、より詳細には、スピーチ信号が発生されて伝達される音響環境から導出したデータでデジタル化されたスピーチ信号を補償することに係る。

【0002】

【従来の技術】来る数年間に、スピーチは、コンピュータシステムと対話するための最も使用される入力方式の1つになることが予想される。キーストローク、マウスクリック及び目に見える身体の身振りに加えて、スピーチは、ユーザがコンピュータ化システムと対話する方法を改善することができる。処理されたスピーチは、我々が何と言ったかを聞き分けそして我々が誰であるかも見出すように認識することができる。スピーチ信号は、コンピュータシステムへのアクセスを得そして音声コマンド及び情報を用いてシステムを動作するように益々利用

される。

【0003】スピーチ信号が「クリーン」であって、音響的に素朴な環境で発生される場合には、良好な結果を生じるための信号の処理作業は、比較的単純である。しかしながら、我々は、システムと対話するための種々様々な異なる環境、例えば、オフィスや、家庭や、道路際の電話や、或いはこれについてはセルラー電話を携帯できるとこの場所でも、スピーチを使用するので、効率的で健全なスピーチ処理を与えるためには、これらの環境における音響的な相違を補償することが重要な問題となる。

【0004】一般に、2つの形式の作用がクリーンスピーチを「ダーティ」にさせる。第1の作用は、スピーチ信号自体の歪である。音響環境は、数えきれないほどの多数の仕方で音声信号を歪ませる。信号は、予想不能に遅延され、進まされ、複製されて、エコーを発生し、周波数及び振幅を変化し、等々である。更に、異なる形式の電話、マイクロホン及び通信ラインは、更に別の異なる歪を導入し得る。

【0005】第2の汚染作用は「ノイズ」である。ノイズは、元々のスピーチの部分ではない付加的な信号がスピーチ周波数スペクトルに生じることによるものである。ノイズは、背後で話をしている他の人、オフィスの装置、自動車、飛行機、風等により導入され得る。通信チャンネルにおける熱的なノイズもスピーチ信号に付加され得る。「ダーティ」スピーチを処理する問題は、歪及びノイズが時間と共に動的に変化することにより更に複雑になる。

【0006】一般に、効率的又は健全なスピーチ処理は、次の段階を含む。第1段階では、デジタル化されたスピーチ信号が時間整列された部分（フレーム）に仕切られ、この場合、直線の予想係数（LPC）「特徴」ベクトルにより音響的特徴を一般に表すことができる。第2段階では、環境的音響データを使用して、ベクトルをクリーンアップすることができる。即ち、ダーティスピーチ信号を表すベクトルに処理を適用し、相当量のノイズ及び歪が除去される。クリーンアップされたベクトルは、統計学的な比較方法を使用して、クリーンな環境で発生された同様にスピーチに厳密に類持される。次いで、第3段階では、クリーンな状態にされた特徴ベクトルは、スピーチがいかに使われようとしているかを決定するスピーチ処理エンジンに送られる。典型的に、この処理は、統計学的モデル又はニューラルネットワークを用いてスピーチ信号パターンを分析及び識別することに依存する。

【0007】別の解決策においては、特徴ベクトルがダーティのままにされる。むしろ、スピーチを処理するのに使用される予め記憶された統計学的モデル又はネットワークは、ダーティスピーチの特徴ベクトルの特性に類似するように変更される。このように、クリーンスピー

チとダーティスピーチとの間、又はそれらの代表的な特徴ベクトルの間の不一致を減少することができる。

【0008】データ、即ち特徴ベクトルではなく、プロセス（又はスピーチ処理エンジン）それ自体に補償を適用することにより、最大化がスピーチ信号及び環境パラメータの両方に及ぶような一般化された最大見込みの問題を解決するためのスピーチ分析を構成することができる。このような一般化されたプロセスは性能を改善するが、計算という点で甚だしいものになる傾向がある。従って、「ダーティ」スピーチ信号のリアルタイム処理を必要とする公知の用途は、プロセスではなくて信号をコンディショニングする傾向が強く、満足な結果をほとんど生じない。

【0009】補償型のスピーチ処理は、近年益々精巧になってきている。初期の処理の幾つかは、ケプストラム平均正規化（CMN）及び相対的スペクトル（RASTA）方法を使用している。これら方法は、同じ平均減算方法の2つの変形である。従って、その考え方は、到来するスピーチフレームから、測定されたスピーチの推定値を減算することである。古典的なCMNは、測定された全てのスピーチを表す平均値を各スピーチフレームから減算するが、RASTAは、平均値の「遅れ」推定値を各フレームから減算する。CMN及びRASTAの両方の方法は、チャンネル特性の相違を直接的に補償し、改善された性能を生じる。両方の方法は、比較的簡単な実施手段を使用するので、多くのスピーチ処理システムに頻繁に使用される。

【0010】第2の種類の効率的な補償方法は、ステレオ記録に依存している。一方の記録は、スピーチ処理システムが既にトレーニングされたところの高性能マイクロホンで行われ、他方の記録は、システムに適應されるべきターゲットマイクロホンで行われる。この解決策は、再トレーニングのためのスピーチ統計情報のブートストラップ推定値を与えるように使用できる。クリーン及びダーティの両スピーチの同時記録をベースとするステレオ対方法は、この問題に対して非常に有用である。

【0011】確率的に最適なフィルタ（POF）方法では、ベクトルコードブック（VQ）が使用される。VQは、コードワード依存の多次元横断フィルタに組み合わされたクリーンスピーチのメル周波数ケプストラム係数（MFCC）の分布を示す。このフィルタの目的は、時間的にずらされたスピーチのフレーム間の時間的相関を得ることである。POFは、予想されるスピーチと測定されたスピーチとの間の最小平方エラー基準の最小化を使用して各フレーム依存VQフィルタ（マトリクス）及び各環境のパラメータを「学習」する。

【0012】POF方法と同様の別の既知の方法である固定コードワード依存ケプストラム正規化（FCDN）も、クリーンスピーチのケプストラムベクトルの分布に対するVQ表示を使用する。この方法は、同時に記

録されたスピーチに基づいてコードワード依存修正ベクトルを計算する。この方法は、その効果として、クリーンスピーチからダーティスピーチへの変換のモデリングを必要としない。しかしながら、この効果を得るために、ステレオ記録が必要とされる。一般に、これらのスピーチ補償方法は、ケプストラムベクトルに対する環境の作用がステレオ記録を用いて直接的にモデリングされるので、環境について何らの仮定も行わない。

【0013】1つの方法であるコードワード依存ケプストラム正規化（CDN）では、クリーンスピーチ信号のケプストラムは、各ガウスをその平均及び共変量で表すことのできるガウス分布の混合体を用いてモデリングされる。CDN方法は、クリーンスピーチケプストラムの分布に対する環境の作用を分析的にモデリングする。この方法の第1段階では、観察されるダーティケプストラムベクトルの見込みを最大にするための環境パラメータ（ノイズ及び歪）の値が推定される。第2段階では、ダーティスピーチのケプストラムベクトルが与えられたときに、クリーンスピーチの観察されないケプストラムベクトルを発見するために、最小平均平方推定（MMSE）が適用される。

【0014】この方法は、通常、センテンスごとに即ちバッチベースで機能し、それ故、環境パラメータを推定するのに非常に長いスピーチサンプル（例えば、2、3秒）を必要とする。バッチ処理により待ち時間が導入されるので、この方法は、連続的なスピーチ信号のリアルタイム処理にはあまり適していない。並列組合せ方法（PMC）は、CDN方法に使用されたものと同じ環境モデルを仮定する。ノイズ及びチャンネル歪ベクトルが完全に分かっていると仮定すれば、この方法は、隠れたマルコフモデル（HMM）の音響分布の平均ベクトル及び共変量マトリクスを変換して、HMMをダーティスピーチのケプストラムの理想的な分布に類似させるように試みる。

【0015】平均ベクトル及び共変量マトリクスを変換するための多数の種々の技術が知られている。しかしながら、PMCのこれら全ての変形は、ノイズ及びチャンネル歪ベクトルを前もって知ることが必要である。推定は、一般に、異なる近似を用いて前もって行われる。通常、分離されたノイズのサンプルは、PMCのパラメータを十分に推定することが必要とされる。これらの方法は、チャンネルの歪が測定されたスピーチ統計情報の平均に影響し、そして特定の周波数における有効なSNRが測定されるスピーチの共変量を制御することを示している。

【0016】スピーチ補償のためのベクトルテイラー級数（VST）方法を用いると、このことを利用して、クリーンスピーチの統計情報が与えられたときにダーティスピーチの統計情報を推定することができる。VTS方法の精度は、テイラー級数近似の上位項のサイズに依存

する。上位項は、スピーチ統計情報の共変量のサイズにより制御される。VTSでは、スピーチは、ガウス分布の混合体を用いてモデリングされる。スピーチを混合体としてモデリングすることにより、各個々のガウスの共変量は、スピーチ全体の共変量より小さくなる。VTSが機能するためには、最大化段階を解決するために混合体のモデルが必要であると示すことができる。これは、パラメータ推定のための十分な潤沢さの概念に関連している。

【0017】

【発明が解決しようとする課題】要約すれば、既知の最良の補償方法は、ガウス分布の混合体におけるクリーンスピーチ特徴ベクトルの確率密度関数 $p(x)$ についてのそれらの表示をベースとする。これらの方法は、パッチモードで機能し、即ち処理を行う前に実質的な量の信号を「聞く」必要がある。これらの方法は、通常、環境パラメータが決定論的であり、それ故、確率密度関数では表されないと仮定する。最後に、これらの方法は、ノイズの共変量を推定するための容易な仕方を与えるものではない。これは、常に収束することが保証されない発見的な方法により共変量を学習しなければならないことを意味する。

【0018】そこで、クリーンスピーチ信号を自然に表すことのできるスピーチ処理システムを提供することが要望される。更に、このシステムは、連続的なスピーチを、それが受け取られたときに、不当な遅延を伴うことなく処理できるように、フィルタとして機能しなければならない。更に、このフィルタは、クリーンスピーチをターンさせる環境パラメータが時間と共にダーティ変化するとき自身を適応させねばならない。

【0019】

【課題を解決するための手段】本発明は、その広い形態において、請求項1に記載するように、歪のないクリーンなスピーチ信号を基準として使用することにより、歪のあるスピーチ信号を処理するためのコンピュータ化された方法に係る。環境ノイズ及び歪パラメータ Q 、 H 及び Σ_n の推定値を使用して連続的なダーティスピーチ信号を補償するためのコンピュータ化された方法が提供される。この方法において、クリーンスピーチ信号を表す第1の特徴ベクトルがベクトルコードブックに記憶される。 Q 、 H 及び Σ_n によりパラメータ化されたノイズ及び歪を含むダーティスピーチ信号に対して第2のベクトルが決定される。

【0020】ノイズ及び歪パラメータは、第2ベクトルから推定される。推定されたパラメータを使用して、第3のベクトルが推定される。第3のベクトルは、第2ベクトルに適用されて、修正されたベクトルを形成し、この修正されたベクトルを第1ベクトルと統計学的に比較して、その修正されたベクトルに最も類似する第1ベクトルを識別することができる。好ましくは、第3のベク

トルは、ベクトルコードブックに記憶することができる。比較の間に、特定の修正されたベクトルと、それに対応する第1ベクトルとの間の距離を決定することができる。この距離は、第1ベクトルが上記修正されたベクトルに類似する見込みを表す。更に、特定の修正されたベクトルがそれに対応する第1ベクトルに類似する見込みが最大にされる。

【0021】スピーチ認識システムにおいては、修正されたベクトルを使用して、ダーティスピーチの発音内容を決定し、スピーチ認識を行うことができる。話し手識別システムにおいては、修正されたベクトルを使用して、ダーティスピーチ信号を発する未知の話し手の認識を決定することができる。本発明の実施形態においては、ノイズ及び歪パラメータが時間と共にダーティスピーチを変化させるときに、第3ベクトルが動的に適応される。

【0022】

【発明の実施の形態】以下、添付図面を参照し、本発明の好ましい実施形態を詳細に説明する。図1は、本発明の好ましい実施形態による適応補償型スピーチ処理システム100の概要を示す。トレーニング段階中に、クリーンスピーチ信号101がマイクロホン（図示せず）により測定される。以下、クリーンスピーチとは、ノイズ及び歪のないスピーチを意味する。

【0023】クリーンスピーチ101は、デジタル化され（102）、測定され（103）そして統計学的にモデリングされる（104）。クリーンスピーチ101を表すモデリング統計情報 $p(x)$ 105は、スピーチ処理エンジン110により使用するためにベクトルコードブック（VQ）106のエントリーとしてメモリに記憶される。トレーニング後に、システム100は、ダーティスピーチ信号を処理するのに使用できる。

【0024】この段階中に、スピーチ信号 $x(t)$ 121は、上記トレーニング段階中に使用されたマイクロホンに対して電力スペクトル $Q(\cdot)$ 122を有するマイクロホンを用いて測定される。実際の使用中に存在する環境条件により、スピーチ $x(t)$ 121は、未知の加算的な静的ノイズ及び未知の直線的なフィルタ作用、例えば、歪 $n(t)$ 123によりダーティ状態にされる。これらの加算的な信号は、電力スペクトル $H(\omega)$ 124をもつフィルタを通過するホワイトノイズとしてモデリングすることができる。

【0025】ノイズ及び歪がここで（125）加算されること、又は信号 $x(t)$ 125がマイクロホンで測定される前に加算されることは、構造的に同等であることに注意されたい。いずれの場合にも、実世界の環境条件は、ダーティスピーチ信号 $z(t)$ 126を生じさせる。ダーティスピーチ信号126は、デジタル信号プロセッサ（DSP）200により処理される。

【0026】図2は、DSP200を詳細に示す。DS

P200は、ダーティ信号 $z(t)$ 126の時間整列された部分を選択し(210)、そしてその部分に良く知られた窓関数、例えば、ハミング窓を乗算する。段階230において、窓処理された部分220に高速フーリエ変換(FFT)が適用され、「フレーム」231が形成される。好ましい実施形態では、選択されたデジタル化部分は、410個のサンプルを含み、これに410ポイントのハミング窓が適用されて、512ポイントのFFTフレーム231が形成される。

【0027】次いで、段階240において、FFT結果の平方の大きさを得ることにより、フレーム231に対する周波数電力スペクトル統計情報が決定される。FFT項の半分は、冗長なものであるから、落とすことができ、256ポイントの電力スペクトル推定値が残される。段階250において、スペクトル推定値は、これにメル周波数の回転マトリクスを乗算することによりメル周波数ドメインへと回転される。段階260は、回転さ

$$z(\omega, T) = \log(\exp(Q(\omega) + x(\omega, T)) + \exp(H(\omega) + n(\omega, T)))$$

但し、 $x(\omega, T)$ は、ノイズ及びチャンネル歪を伴わずに測定された基礎となるクリーンベクトルであり、そして $n(\omega, T)$ は、ノイズ及び歪のみが存在した場合の統計情報である。

【0030】ノイズのない状態では、チャンネルの電力スペクトル $Q(\omega)$ 122が、測定信号 $x(t)$ 121に直線的な歪を発生する。ノイズ $n(t)$ 123は、電力スペクトルドメインにおいて直線的に歪まされるが、対数スペクトルドメインでは非直線的である。更に、 E

$$\Sigma_z = \text{diag}(b/b+1))\Sigma_x \text{diag}(b/b+1)) + \text{diag}(1/b+1))\Sigma_n \text{diag}(1/b+1))$$

を発生することにより、決定することができる。ここで、周波数及び時間に対する項の依存性は明瞭化のため

$$b = \exp(Q + E[x] - H - E[n])$$

【0032】式2及び3は、チャンネルが、測定された統計学的情報の平均を直線的にシフトし、信号対雑音比を減少し、そしてノイズの共変量がスピーチの共変量より小さいので測定されたスピーチの共変量を減少することを示している。この分析に基づき、本発明は、上記したVTS及びPMCの公知方法を独特に結合して、ダーティスピーチの動的に変化する環境パラメータに適應する補償型スピーチ処理方法を可能にする。

【0033】本発明は、トレーニングスピーチを環境補償の目的でベクトル $p(x)$ としてそれ自体で自然に表すことができるという考え方を使用する。従って、全てのスピーチは、トレーニングスピーチベクトルコードブック(VQ)107により表される。加えて、クリーンなトレーニングスピーチと、実際のダーティスピーチとの間の差は、予想最大化(EM)プロセスを用いて決定される。以下に述べるEMプロセスでは、予想段階と最大化段階が繰り返し実行されて、勾配上昇中に最適な結

れた推定値の対数を取り、各フレーム231に対する特徴ベクトル表示261が得られる。

【0028】段階270の更に別の考えられる処理は、メル周波数の対数スペクトルに離散的コサイン変換(DCT)を適用してメルケプストラムを決定することを含む。メル周波数変換は任意であり、これを伴わないDCTの結果は、単にケプストラムと称する。処理中に、窓関数は、測定されたダーティ信号 $z(t)$ 126に沿って移動する。DSPの段階200は、ハミング窓の各新たな位置において信号に適用される。その正味の結果は、特徴ベクトル $z(\omega, T)$ 128のシーケンスである。このベクトル128は、図1のエンジン110により処理することができる。このベクトル128は、VQ107のエントリーと統計学的に比較され、結果199が得られる。

【0029】ノイズ及びチャンネル歪は、ベクトル128に次のように作用することが示される。

式1

エンジン110は、 $x(\omega, T)$ の統計学的表示、例えば、VQ107にアクセスすることに注意されたい。本発明は、この情報を用いて、ノイズ及び歪を推定する。

【0031】スピーチ統計情報に対するノイズ及び歪の作用は、次の一次テイラー級数拡張

$$E[z] = Q + E[x] + \log(1 + 1/b)$$

を用いて、クリーンスピーチベクトルの平均値に対して式1を拡張し、

式2

に落としてある。これは、歪の作用が信号対雑音比に依存し、これは、次のように表すことができる。

式3

果に向かって収斂させる。記憶されたトレーニングスピーチ $p(x)$ 105は、数1のように表すことができる。

【0034】

【数1】

$$p(x) = \sum_i P_i \delta(x - v_i)$$

【0035】但し、集合 $\{V_i\}$ は、全ての考えられるスピーチベクトルに対するコードブックを表し、そして P_i は、対応するベクトルによりスピーチが発生された以前の確率である。

【0036】この表示は、コードブックのサイズが非常に大きなものでない限り、スピーチの認識には適当でないが、健全なパラメータの推定及び補償のための優れた表示である。これが真である理由は、健全なスピーチ処理システムは、EMプロセスを用いて分布から推定できるある全体的なパラメータの統計情報を推定するだけで

よいからである。

【0037】図3に示すように、補償プロセス300は、3つの主たる段階を含む。EMプロセスを用いる第1段階310において、ノイズ及び（チャンネル）歪のパラメータが決定され、これらパラメータがベクトルコードブック107に送られたときに、コードブックは、変換されたコードブックがダーティスピーチを最良に表す見込みを最大にする。EMプロセスが収斂した後の第2段階320において、推定された環境パラメータが与えられると、コードブックベクトル107の変換を予想する。この変換は、1組の修正ベクトルとして表すことができる。

【0038】第3段階330の間に、修正されたベクトルが、到来するダーティスピーチの特徴ベクトル128に付与され、それらを、最小平均平方エラー（MMS

$$V'_i \leftarrow \log(\exp(Q + V_i) + \exp(H))$$

ここで、値 $E[n]$ は、 H の値に吸収されている。ノイズに対するこの関係の第1導関数は、数2の通りである。

【0040】

【数2】

$$F_i(i, j) = \delta(i - j) \frac{\exp(H_i)}{\exp(Q_i + x_i)}$$

【0041】但し、 $\delta(i - j)$ は、クロンカーデルタである。

【0042】各予想されたコードワードベクトル V'_i は、次いで、数3のように変換される以前のものにより拡張される（420）。

【0043】

【数3】

$$\sqrt{-1/2 \log(P_i)}$$

【0044】又、各ダーティスピーチベクトルは、ゼロにより増大される（430）。このように、増大されたダーティベクトルと、増大された V'_i コードワードを直接比較することができる。完全に拡張されたベクトル V'_i は、数4で表される。

【0045】

【数4】

$$\begin{bmatrix} V_i \\ \sqrt{-1/2 \log(P_i)} \end{bmatrix}$$

【0046】そして増大されたダーティベクトルは、数5の式を有する。

【0047】

【数5】

E) という意味で、VQ107に記憶されたクリーンベクトルに類似させる。1つの効果として、本発明の補償プロセス300は、処理エンジン110とは独立しており、即ち補償プロセスは、ダーティ特徴ベクトルに対して動作して、ベクトルを修正し、環境におけるノイズ及び歪により汚染されていないクリーンスピーチから導出されたベクトルにそれらが密接に類似するようにする。

【0039】これら段階の細部を詳細に説明する。図4に示すように、EM段階は、環境を特定する3つのパラメータ $\{Q, H, \Sigma_n\}$ を繰り返し決定する。第1段階410は、予想段階である。 $\{Q, H, \Sigma_n\}$ の現在値は、コードブック107の各ベクトルを、各々式1を用いて予想された修正ベクトル V'_i へとマップするのに使用される。

$$z'_i = \begin{bmatrix} z_i \\ 0 \end{bmatrix}, \quad \text{式4}$$

【0048】これにより得られる1組の拡張された修正ベクトルは、次いで、ベクトルコードブックVQに記憶することができる（440）。例えば、コードブックの各エントリは、音響環境の現在状態を反映する現在関連する拡張された修正ベクトルを有することができる。この拡張された修正ベクトルは、コードブックベクトルと、対応するダーティスピーチベクトル128との間の距離の $-1/2$ 倍を、ダーティベクトル z_t がコードワードベクトル v_i で表される見込みとして使用できるという特性を有する。

【0049】図5は、予想段階500を詳細に示す。この段階中に、到来するダーティベクトル128の1つと、（修正された）コードブックベクトルとの間の最良の一致が決定され、そして最大化段階に必要な統計情報が累積される。プロセスは、段階501において、変数 L, N, n, Q, A 及び B を0に初期化することにより始まる。図5に示すように、各到来ダーティベクトル128について、次の段階が実行される。まず、段階502において、変換されたベクトルに最も類似する新たなベクトルコードブックのエントリ $VQ(z^e)$ を決定する。クリーンベクトルに関連したコードブックの初期修正ベクトルは、0にすることもできるし、推定することもできる点に注意されたい。このエントリへのインデックスは、次のように表される。

$$j(i) = \arg \min [k] | VQ(z^e_k), [z'_t, 0] |^2$$

【0051】更に、最良のコードブックベクトルと到来するベクトルとの間の平方距離 $(d(z'_i))$ は、段階503において戻される。この距離、即ち選択されたコードブックベクトルとダーティベクトルとの間の統計学的な差は、測定されたベクトルの見込みを次のように

決定するのに使用される。

$$l(z_i) \leftarrow 1/2 d(z'_i)$$

上記のように、これにより得られる見込みは、測定されたダーティベクトルが実際にコードブックベクトルにより表されるその後の確率であることに注意されたい。次いで、見込み $l(z_i)$ は、 $L = L + 1(z_i)$ のように累積され(504)、残留する v_i が段階505において決定される。段階506では、その残留物がガウス分布でホワイト化される。

【0052】次いで、残留物と、ノイズに対する第1導関数との積 $\alpha \leftarrow F(j(i))v$ を計算する(507)。この演算は、 $F(j(i))$ が対角マトリクスであるのでポイントごとの乗算を用いて行うことができる。これに続いて、平均の比を決定する(508)。但し、 $r_1 = n/(n+1)$ 及び $r_2 = 1/(n+1)$ である。ここで、 n は、繰り返し中にそれまで使用された測定されたベクトルの全数である。段階507で決定さ

$$\begin{bmatrix} \sum_n -B & -B^T & +A \\ & & \\ & & -A & +B^T \end{bmatrix} + \sum_q \begin{bmatrix} -A & +B \\ & \\ A & + \sum_N \end{bmatrix} \delta = \begin{bmatrix} Q_i \\ N_i \end{bmatrix}$$

【0056】但し、 Σq 及び ΣN は、 Q 及び N パラメータに指定された以前の共変量を表す。これにより得られた値は、次いで、環境パラメータの現在の推定値に加えられる。EMプロセスが収斂した後に（これは見込みを監視することにより決定できる）、所望のスピーチ処理用途に基づいて最終的な2つの段階を行うことができる。第1段階は、EMプロセスからの環境の推定パラメータが与えられたときにダーティスピーチの統計学的情報を予想する。これは、EMプロセスの予想段階と同等である。第2段階は、その予想された統計学的情報を使用して、MMSE修正ファクタを推定する。

【0057】スピーチ認識

図6に示すように、環境的に補償されたスピーチを使用できる第1の用途は、スピーチ認識エンジンである。ここでは、何が言われたかを決定することが所望される。この用途は、平易な古い電話サービス(POTS)の場合よりもノイズ及び歪が大きくなる傾向のあるセルラー電話ネットワークにわたって収集されたスピーチを認識するのに有用である。又、この用途は、多数の異なる形式のハードウェアシステム及び通信ラインを用いて全世界中の環境においてスピーチを発生することのできるワールドワイドウェブにわたって収集されたスピーチに使用することもできる。

【0058】図6に示すように、ダーティスピーチ信号601は、デジタル化処理され(610)、ダーティ特徴ベクトルの時間的シーケンス602を発生する。各ベクトルは、連続スピーチ信号のセグメントに見られる1

れた積は、段階509で累積される。段階509の積と残留物との間の差は、段階510において、次のように累積される。

$$Q_s \leftarrow r_1 Q_s + r_2 (V^* i - \cdot)$$

次いで、段階511において、ノイズの共変量が推定し直される。最後に、段階512において、変数 A が次のように累積される。

$$[0053] A \leftarrow r_1 A + r_2 (F_1(j(i))^T \Sigma_n^{-1} F_1(j(i)))$$

そして変数 B は、次のようにされる。

$$B \leftarrow r_1 B + r_2 \Sigma_n^{-1} F_1(j(i))$$

【0054】現在推定繰り返しの累積された変数は、次いで、最大化段階に使用される。この最大化は、数6の線型方程式の組を解くことを含む。

【0055】

【数6】

組の音響特徴を統計学的に表す。段階620において、ダーティベクトルは、上記のようにクリーンな状態にされ、「クリーン」ベクトル603を発生する。即ち、本発明を使用し、環境がダーティベクトルに及ぼす影響を取り去る。ここで処理されるべきスピーチ信号は、連続的であることに注意されたい。スピーチの短いバーストに対して動作するバッチ式のスピーチ処理とは異なり、ここでは、補償プロセスは、フィルタとして振る舞う必要がある。

【0059】スピーチ認識エンジン630は、既知の音素605を表す一連の考えられる統計学的パラメータに対しクリーンなベクトル603を一致させる。この一致は、音素シーケンスの多数の考えられる仮説を探索するビタビデコードのような最適なサーチアルゴリズムを用いて効率的に行うことができる。観察されたベクトルのシーケンスに統計学的な意味で最も近い音素の仮説シーケンスが、発音されたスピーチとして選択される。

【0060】図7に示すように、スピーチ認識についてここに述べる補償を使用すると、音声分類作業として背景ノイズに対する健全さが高められる。図7において、 y 軸701は、正しいスピーチと仮説するときの精度%を示し、 x 軸702は、相対的なノイズレベル(SNR)を示す。破線の曲線710は、補償されないスピーチ認識の場合であり、そして実線の曲線720は、補償されたスピーチ認識の場合である。明らかなように、オフィス環境について典型的である約25dBより低い全てのSNRにおいて著しい改善が得られる。

【0061】話し手の確認

図8に示す用途では、話し手が何を話すかは独立して、話し手が誰であるかを決定することが望まれる。ここでは、未知の話し手のダーティスピーチ信号801が処理されて、ベクトル810が抽出される。このベクトル810は、補償されて(820)、クリーンなベクトル803を発生する。このベクトル803は、既知の話し手のモデル805に対して比較され、識別(ID)804が発生される。モデル805は、トレーニングセッションの間に収集できる。

【0062】ここでも、上記と同様に、予想最大化段階で推定された環境パラメータの値が与えられたときに、ノイズのあるスピーチの統計学的情報が最初に予想される。次いで、その予想された統計学的情報が最終的な統計学的情報へとマップされ、スピーチに対して必要な処理が行われる。多数の考えられる技術を使用することができる。1つの技術においては、予想される統計学的情報に対して平均値及び共変量が決定される。次いで、特定の話し手により任意の発音が発せられた見込みを、演算高調波球状度(AHS)又は最大見込み(ML)距離として測定することができる。

【0063】別の考えられる技術は、EMプロセスにより決定された見込みを使用する。この場合には、EMプロセスの収斂後に、それ以上の計算は不要である。図9に示すように、EMプロセスは、ML距離を使用するよりも良好な結果を与えることが実験により示唆される。図9において、y軸901は、話し手を正しく識別する精度%であり、そしてx軸は、SNRの異なるレベルを示す。曲線910は、クリーンスピーチでトレーニングされたモデルと、ML距離計測とを使用する補償されないスピーチの場合である。曲線920は、所与の測定されたSNRにおける補償されたスピーチの場合である。家庭やオフィスにおいて通常見られるSNRが25dB未満の環境では、著しい改善が得られる。

【0064】以上、本発明の特定の実施形態を詳細に説

明した。しかしながら、上記実施形態を変更しても、本発明の効果の幾つか又は全部が達成され得ることは当業者に明らかであろう。従って、このような変更は、全て、本発明の範囲内の包含されるものとする。

【図面の簡単な説明】

【図1】本発明の実施形態によるスピーチ処理システムの流れ線図である。

【図2】連続的なスピーチ信号から特徴ベクトルを抽出するプロセスを示す流れ線図である。

【図3】推定値最大化プロセスの流れ線図である。

【図4】ベクトルを予想するための流れ線図である。

【図5】ベクトル間の差を決定するための流れ線図である。

【図6】スピーチを認識するプロセスの流れ線図である。

【図7】スピーチ認識方法の精度を比較するグラフである。

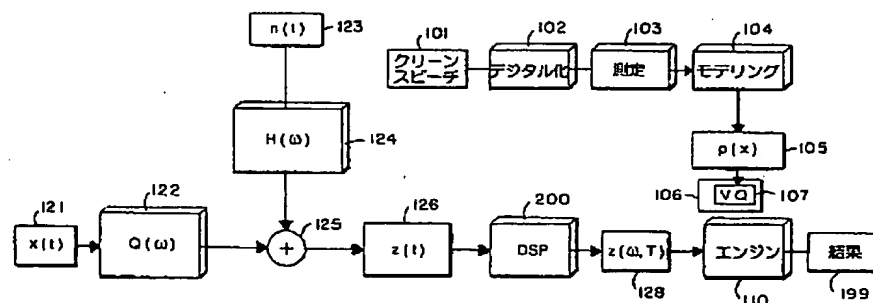
【図8】話し手を確認するプロセスの流れ線図である。

【図9】話し手を確認する方法の精度を比較するグラフである。

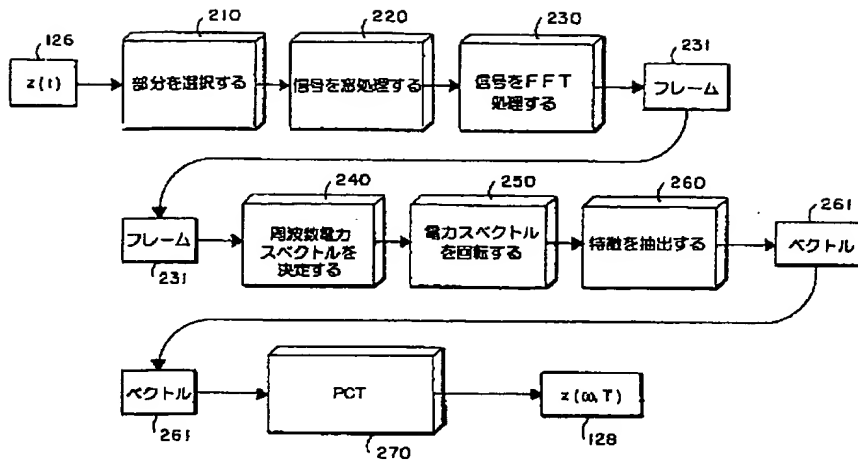
【符号の説明】

- 100 適応補償型スピーチ処理システム
- 101 クリーンスピーチ
- 102 デジタル化
- 103 測定
- 104 モデリング
- 106 ベクトルコードブック
- 110 スピーチ処理エンジン
- 121 スピーチ信号
- 122 電力スペクトル
- 123 歪
- 124 電力スペクトル
- 126 ダーティスピーチ信号
- 200 デジタル信号プロセッサ
- 231 フレーム

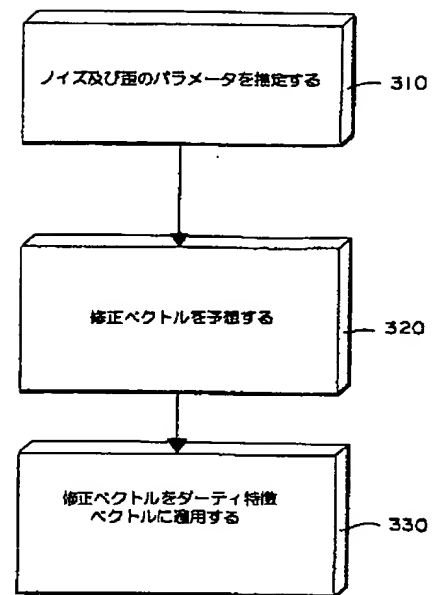
【図1】



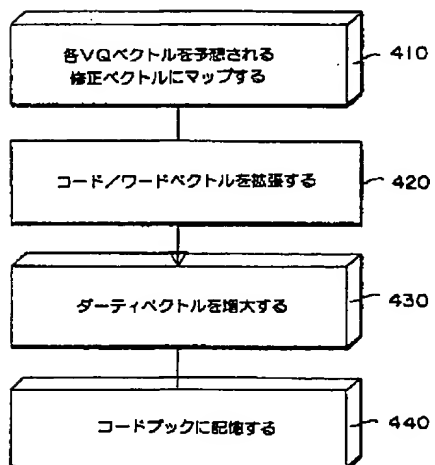
【図2】



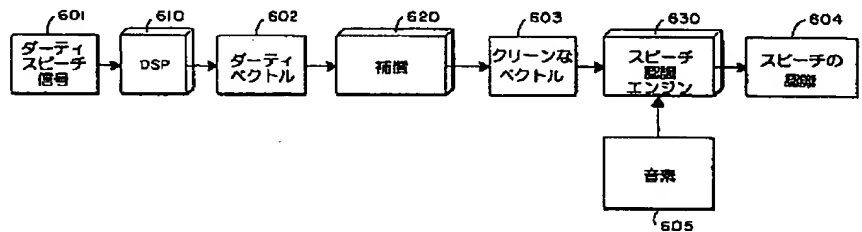
【図3】



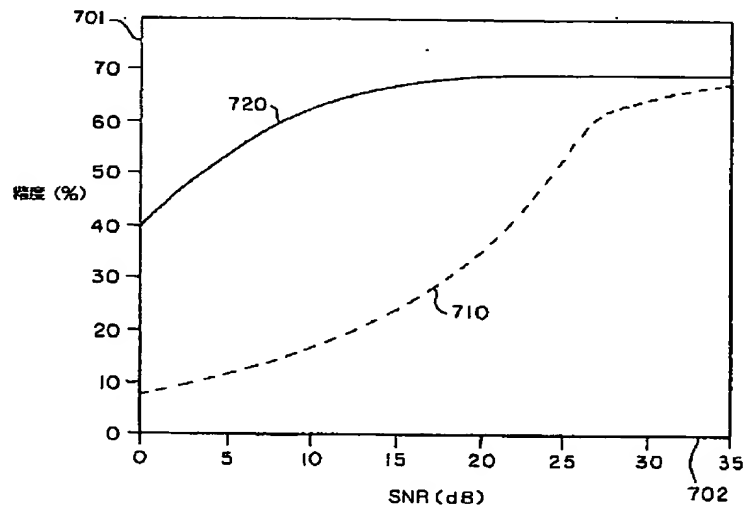
【図4】



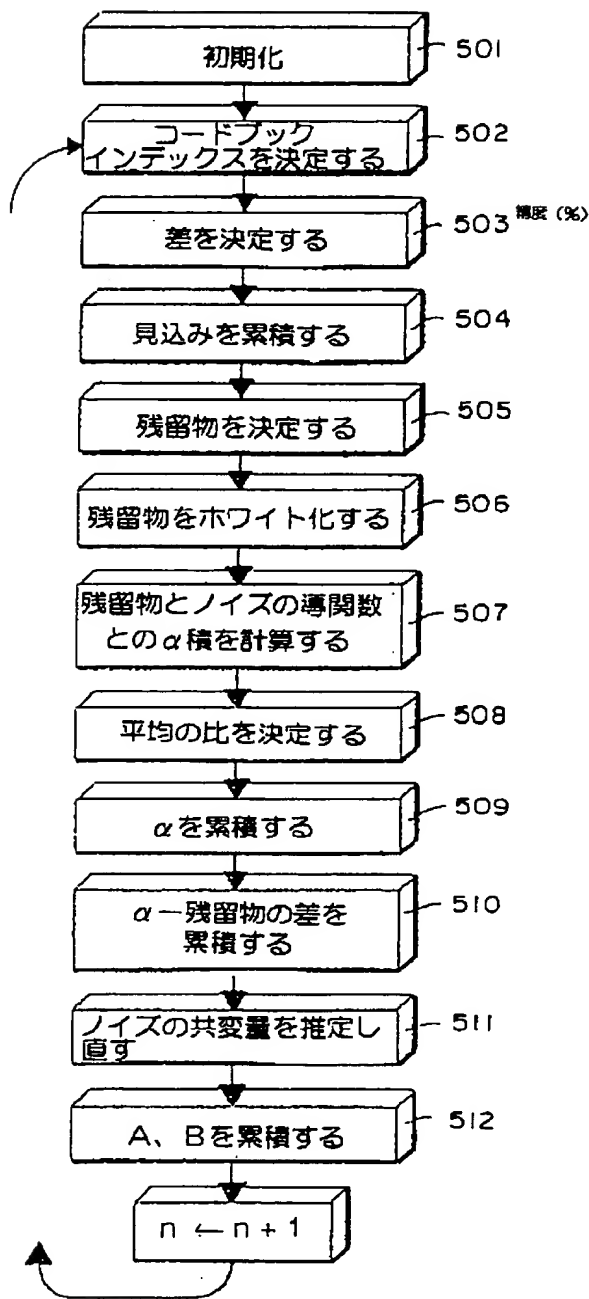
【図6】



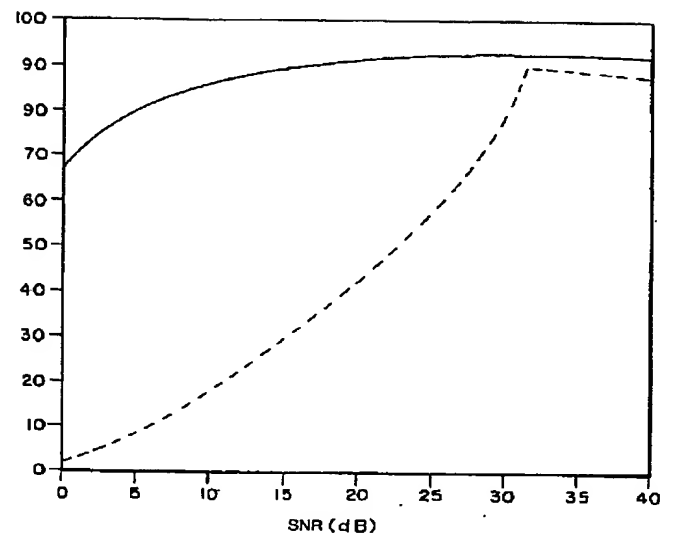
【図7】



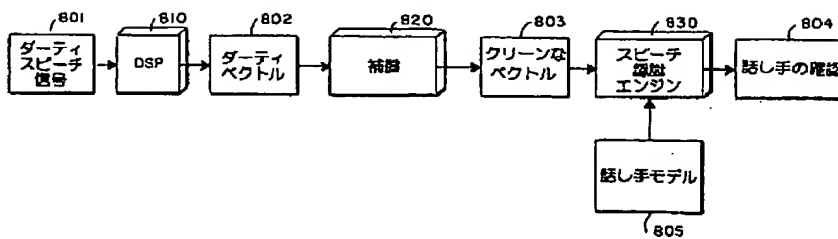
【図5】



【図9】



【図8】



フロントページの続き

(72)発明者 ベドロ ジェイ モレノー
アメリカ合衆国 マサチューセッツ州
02139ケンブリッジ フランクリン スト
リート 345-505